



Background and Methodological Note for the Spatial Education Inequalities*

SEI website

www.spatial-education-inequalities.com

Marcos Delprato^{a,*}

^a*Centre for International Education, University of Sussex, Brighton UK*

Abstract

This note provides a motivation for the SEI website as well as some of the technical details on spatial analysis and computational issues needed for the website. Moreover, the note contains guidelines for users of the website.

*I would like to thank the *ESW impact fund* from the School of Education and Social Work (University of Sussex) for the financial assistance with the production of the SEI website. Furthermore, the calculation of the different education indicators was made possible by the publicly available Demographic and Health Surveys (DHS) dataset and linked geographical GPS dataset. I acknowledge the importance of the DHS team for their work and also to the different country statistical offices which made possible the production of the SEI website.

*Corresponding author

Email address: m.delprato@sussex.ac.uk (Marcos Delprato)

1. Motivation of the SEI website

Across the new Sustainable Development Goals (SDG)¹ there is an emphasis on equity. Specifically, for the sustainable development goal for education 4 (SDG4),² the established markers of disadvantage used for monitoring inequality are wealth, gender and location and some of their overlaps.³ In particular, within a given country, educational indicators defined for rural and urban locations are the pillars to assess geographical disadvantages because they supposed to offer a first glance of rural constraints on educational supply and related weakness of labour markets, as well as a wealth effects of education uptake within large and unequal urban areas.

However, in the global south, the urban-rural division is a blurred boundary as pockets of disadvantage can persist within each location type but with a large degree of variation on the extent of inequality. Hence, a more refined geographic measurement of inequality is needed to target efficiently educational inequalities at a more granular level. This is important as correlations of education indicators across the life course based on smaller areas definitions (beyond what rural and rural categorisations can offer) depends on what happens in proximal areas, which may have positive or negative influences on surrounding areas through cluster externalities.

Yet this type of inequality analysis is absent from an education perspective, thereby hindering a robust monitoring of SDG4. The **SEI website** was created to fill this gap. It relies on tools from geography and spatial econometrics to analyse spatial education. The idea underpinning the SEI website is to engage a body of research across SDGs from the angle of SDG4 from a spatial perspective through providing several maps for education indicators for each country, ranging from local correlation, hot and cold spots, and smoothing maps with linked files for local stakeholders further usage.

Specifically, in current education inequalities (EI) tools:

- There is a lack of comparable country or cross-country XY geo dataset on EI.
- Within country, EI data available, at most, is mapped by sub-regions or by urban-rural communities divide.
- Given the lack of an EI spatial platform → it is not possible to monitor education SDG4 (and crossover with other SDGs like the SG5 on gender or empowerment) from a spatial perspective.

Whereas the SEI website:

- Using *geodata* localised indicators allow (with maps and datasets) a more efficient targeting of educational inequalities.

¹See: <https://sustainabledevelopment.un.org>.

²For target details, see: <https://en.unesco.org/education2030-sdg4/targets>

³See, for instance, UNESCO-UIS datasets: <http://data.uis.unesco.org/>, <https://tellmaps.com/sdg4/> and <https://www.education-inequalities.org/>.

- A *visual granular inspection* of data permits focal analysis on undefined units (beyond admin areas), where clusters of inequalities or externalities may exist for a given country and indicator.
- Website aims
 1. Foster spatial analysis widely on EI at a local level through offering maps across EI on the lifecourse, though importantly, files for further exploration with procedures on spatial analysis are included.
 2. Engagement with local statistical and education stakeholder in the global south and geography departments
 3. Boost the use of additional specific-country EI data and push for their inclusion on International SDGs monitoring agenda.

2. Methodological background

2.1. Introduction

The first law of geography is: “*Everything is related to everything else, but near things are more related than distant things*”.⁴ Spatial dependence is common on social science, albeit often ignored. Spatial dependence exists when the expect utility of a unit is influenced by the choices of other units and it can result from:

- coercion, competition, externalities, learning, emulation.
- when agents/unit strategies are partly dependent of strategies chosen by other agents need to account for spatial dependence.
- yet it can also exist when researchers are not directly interested in studying.

There are two forms of spatial dependence: nomadic (where unit of analysis is a single unit/agent) or dyadic (unit is a pair representing the interaction between two units). The **SEI website** assumes nomadic data. Under this type of spatial dependence, the standard models are shown in Table 1:

Table 1: Alternative spatial models

| | |
|----|-------------------------------------------------------------------------------------------------------------------------------------------|
| M1 | Spatial lag (autoregressive) model the Ys in other units of analysis exerts an influence on the Y in the unit under observation |
| M2 | Spatial X model some Xs of other units affects the Y in the unit under observation |
| M3 | Spatial error model the error processes are systematically correlated across units of observation |
| M4 | Spatial dependence combined |

⁴Tobler, W., (1970). A computer movie simulating urban growth in the Detroit region. Economic geography 46 (sup1), 234-240.

2.2. Model

2.2.1. Notation

Before introducing formally the array of spatial models, some notation is needed.

i area (observation), from 1 to N (= community)

y_i dependent (outcome) variable in area i (education indicator)

$x_{i,1}$ 1st independent variable in area i

.

$x_{i,j}$ j th independent variable in area i

.

$x_{i,k}$ last independent variable in area i

ϵ_i error (residual) in area i

Spatial models or Spatial Autoregressive (SAR) models extend linear regression by outcomes in one area to be affected by:

1. Outcomes in nearby areas (=spatial lags of the outcome variable).
2. Covariates from nearby areas (=spatial lags of covariates).
3. Errors from nearby areas (=spatially autoregressive errors).

2.2.2. Specification

The aim of spatial regression is to estimate the relationship between an outcome variable of interest Y and predictors X , taking into account the spatial dependence among observations. A general model (matrix form) includes influences from Y s, X s, and u :

$$y = \lambda W y + X\beta + W X\gamma + u \quad (1)$$

$$u = \rho W u + \varepsilon \quad (2)$$

For different restrictions, different models are obtained:

- $\gamma = 0, \rho = 0$ and $\lambda \neq 0 \rightarrow$ SLM
- $\gamma = 0, \rho \neq 0$ and $\lambda = 0 \rightarrow$ SEM
- $\gamma = 0, \rho \neq 0$ and $\lambda \neq 0 \rightarrow$ SARAR
- $\gamma \neq 0, \rho = 0$ and $\lambda \neq 0 \rightarrow$ SDM

In non-matrix form, there are three leading models:

$$y_i = \alpha + \lambda \sum_k w_{ik} y_k + \beta X_i + \epsilon_i \quad \text{Spatial lag model (3)}$$

$$y_i = \alpha + \gamma \sum_k w_{ik} x_k + \beta X_i + \epsilon_i \quad \text{Spatial X model (4)}$$

$$y_i = \alpha + \beta X_i + \epsilon_i + \rho \sum_k w_{ik} u_k \quad \text{Spatial error model (5)}$$

- The spatial parameter λ gives the impact of the spatial effect variable $\sum_k w_{ik} y_k$.
- When fitting SLM one needs to deal with **endogeneity**: when unit k affects unit i , also unit i affects unit k : $y_k \rightarrow y_i \rightarrow y_k \rightarrow \dots$
- In SXM also: $x_j \rightarrow y_i \rightarrow x_i \rightarrow y_j \rightarrow x_j \rightarrow \dots$
- Need to control for endogeneity \rightarrow generalized spatial two-stage least-squares estimator (**G2SL**).

2.2.3. Weighting matrix

- The spatial matrix **W** captures who is neighbourhood of whom - the relationship among units (exogenous) or spatial structure.
- The building of the W matrix is given by the researcher. The choice of weighting matrix should be based on research question.
- Criteria for W construction can be geographical, socio-economic or a combination.
- We focus on the geographical criteria. Within that, we focus in distance functions.
 - Distance: distance with thresholds \rightarrow we follow this criteria.
 - Contiguity: rook, queen.
 - K nearest neighbours.

The spatial matrix **W** is a symmetric matrix with diagonal elements taking the value of 0, and each row is row standardised. Each element (i, j) specifies the potential spillovers from area i to area j .

$$W = \begin{pmatrix} 0 & w_{1,2} & w_{1,3} & \dots & w_{1,n} \\ w_{2,1} & 0 & w_{2,3} & \dots & w_{2,n} \\ w_{3,1} & w_{3,2} & 0 & \dots & w_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{n,1} & w_{n,2} & w_{n,3} & \dots & 0 \end{pmatrix}$$

The law is \rightarrow spillovers are decreasing with distance. Thus, elements equal to the reciprocal of distance between places (before normalization) based on units (communities) XY coordinates. Functions for w_{ij} :

$$w_{ij} = \frac{d_{ij}^{-\delta}}{\sum_{j=1}^n d_{ij}^{-\delta}} \rightarrow \text{power functional type} \quad (6)$$

$$w_{ij} = \frac{\exp(\delta d_{ij})}{\sum_{j=1}^n \exp(\delta d_{ij})} \rightarrow \text{exponential type} \quad (7)$$

if $d_{ij} < d, i \neq j, \delta > 0$ and 0 otherwise, where δ is distance decay parameter and d distance threshold.

2.3. Spatial autocorrelation

The 1st law of geography (that is: everything is related to everything else, but near things are more related than distant things) defines the statistical concept of (positive) **spatial autocorrelation**, according to which two or more units that are spatially close tend to be more similar to each other (with respect to a given attribute Y) than spatially distant objects. In general, spatial autocorrelation implies **spatial clustering**, that is, the existence of sub-areas of the study area where the attribute of interest Y takes higher than average values (hot spots) or lower than average values (cold spots).

There are two broad categories of spatial autocorrelation:

1. A **global** index of spatial autocorrelation expresses the overall degree of similarity between spatially close regions observed in a given study area A with respect to a variable Y . It detects a general tendency of clustering.
2. A **local** index of spatial autocorrelation expresses, for each region r_i of a given study area A , the degree of similarity between that region and its neighboring regions with respect to a numeric variable Y .

2.3.1. Moran I test

Spatial autocorrelation is key on spatial statistical analysis. As stated, there are two classes.

- **Global** - which measures the extent to which units are generally interdependent. Moran I lies within the range $[-1,1]$. Random distribution of y in space the indicator is near 0. Positive (negative) values of I indicates units neighboring a unit with high (low) values also show high (low) values and test global correlation through $z[I]$.

$$\text{Moran } I = \frac{n}{S_0} \frac{\sum_i \sum_j (y_i - \bar{y}) w_{ij} (y_j - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (8)$$

- **Local** - captures spots showing high spatial autocorrelation locally, detecting special clusters in local dimension.

$$\text{Local Moran } I_i(d) = \frac{(y_i - \bar{y})}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \sum_{j \neq i}^n w_{ij}(d)(y_i - \bar{y}) \quad (9)$$

where $w_{ij}(d)$ is the weighting distance. Test $z[I_i]$ allows to group observations in 4 categories: HighHigh (H-H), LowLow (L-L), LowHigh (L-H), and HighLow (H-L).

2.4. Prediction, data interpolation

- To obtain smoothed country maps based on XY data points for an education indicator, we employ Empirical Bayesian kriging (**EBK**).⁵
- EBK is a geostatistical interpolation method that automates the most difficult aspects of building a valid kriging model.
- Also, EBK allows a better calculation of standard errors of prediction by considering the uncertainty of semivariogram estimation into account through a Bayesian approach, important for smaller datasets.
- Kriging is a probabilistic predictor and have standard errors that quantify the uncertainty associated with the predicted values. Kriging predictors are called optimal predictors because the prediction error is minimized.
- Kriging uses a semivariogram (a function of the distance and direction separating two locations) to quantify the spatial dependence in the data. A semivariogram is constructed by calculating half the average squared difference of the values of all the pairs of measurements at locations separated by a given distance h .
- Smoothing maps can be obtained as well as standard deviation of prediction (alongside statistics like RMSE, to judge the validity of the interpolation) with minimal interaction.

3. SEI website content

3.1. Overview

In the SEI website, <http://spatial-education-inequalities.com/>, users can find an array of information helping to identify education inequality from a spatial perspective. Specifically, users can compare results across countries, though the main

⁵For details, see the following links on the ArcGis software site:

<https://www.esri.com/news/arcuser/1012/empirical-byesian-kriging.html>
<http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/geostatistical-analyst/what-is-empirical-bayesian-kriging-.htm>

purpose of the site is on country-specific analysis. The SEI website contains maps, spatial statistics, and raw datasets and geographical files for further analysis.

Thus far the SEI website covers the sub-Saharan Africa (SSA) region, and it measures spatial education inequality for 29 (latest) SSA DHS samples/countries for $\approx 16,090$ communities.

There are 14 education indicators included in the SEI website. The list of surveys/years for the SSA countries included is shown in Table 2.

Table 2: Sub-Saharan Africa’s DHS surveys used

| Country | Year | Number of communities | SSA region |
|----------------|------|-----------------------|------------|
| Angola | 2015 | 625 | MA |
| Benin | 2011 | 746 | WA |
| Burkina Faso | 2010 | 542 | WA |
| Burundi | 2016 | 552 | EA |
| Cameroon | 2011 | 577 | MA |
| Chad | 2014 | 624 | MA |
| Côte d’Ivoire | 2011 | 341 | WA |
| D. R. Congo | 2013 | 493 | MA |
| Ethiopia | 2016 | 622 | EA |
| Gabon | 2012 | 332 | MA |
| Ghana | 2014 | 423 | WA |
| Kenya | 2014 | 1,585 | EA |
| Lesotho | 2014 | 399 | SA |
| Liberia | 2013 | 322 | WA |
| Madagascar | 2008 | 585 | EA |
| Malawi | 2015 | 850 | EA |
| Mali | 2012 | 413 | WA |
| Mozambique | 2011 | 609 | EA |
| Namibia | 2013 | 550 | SA |
| Nigeria | 2013 | 889 | WA |
| Rwanda | 2014 | 492 | EA |
| Senegal | 2014 | 197 | WA |
| Sierra Leone | 2013 | 435 | WA |
| Swaziland | 2006 | 270 | SA |
| Togo | 2013 | 330 | WA |
| Uganda | 2011 | 608 | EA |
| U. R. Tanzania | 2015 | 400 | EA |
| Zambia | 2013 | 720 | EA |
| Zimbabwe | 2015 | 400 | EA |

For each country of the SSA region, maps are provided (see example in Figure 1).

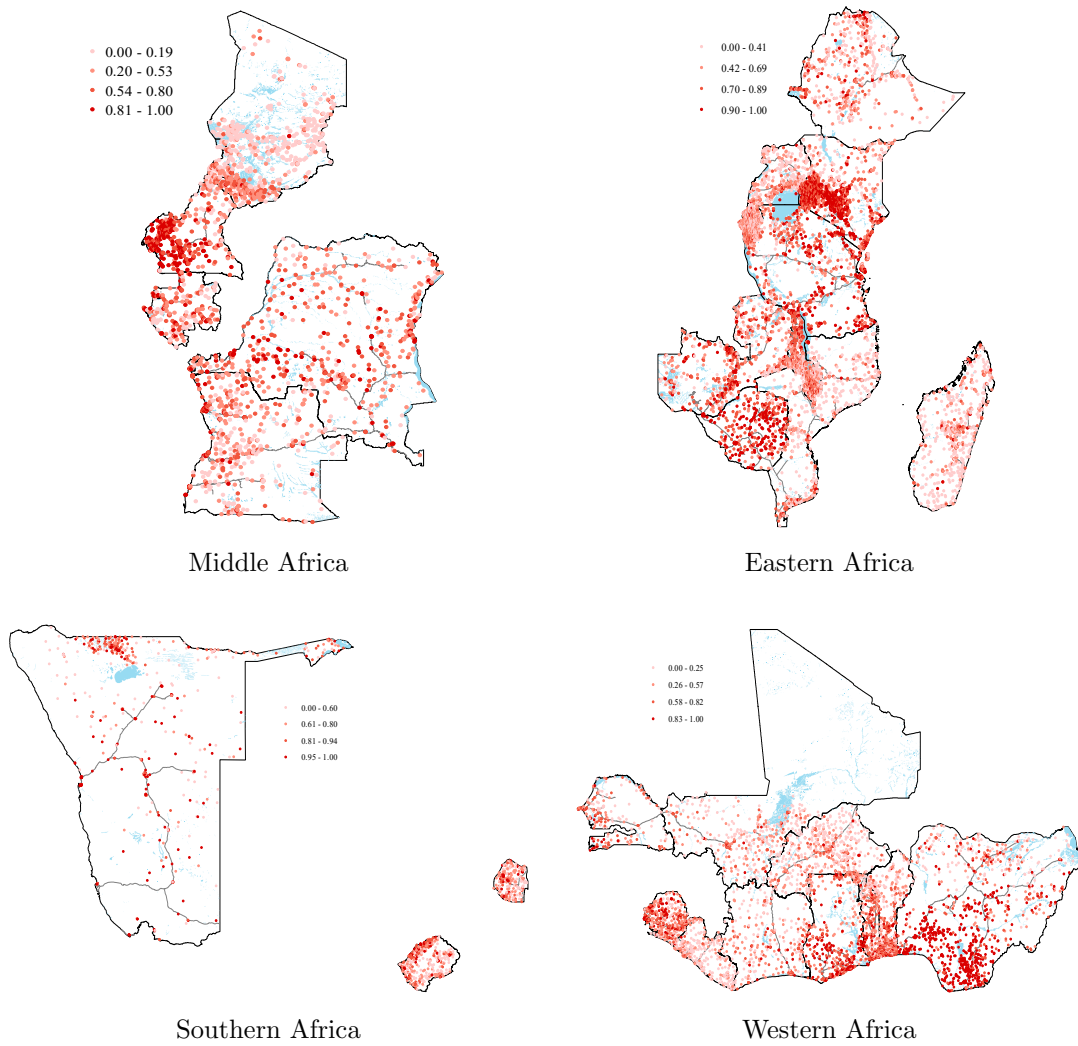


Figure 1: Example. Primary completion - SSA regions

3.2. Data

There are two sources of data within DHS which are required to produce the information of the SEI website.

3.2.1. Standard variables

- The final working dataset is produced by pooling and recoding different datasets (within Measure DHS and others sources).
- Background health indicators are in BR (birth recode) and KR (children dataset).
- Empowerment indicators, early marriage, demographic, etc are from IR (women dataset).
- Educational indicators (=14) and key background socio-economic characteristics (wealth, mother education, work situation, etc.) are from PR (household recode) dataset.
- All variables are averaged at the community level (the unit of analysis).

3.2.2. Geographical variables

- XY coordinates are from Measure DHS dataset folders, **shape files:shp**. It also contains info on admin boundaries and altitude. Note: random displacement of XY.
- Surface **shape files (shp)** on country, admin borders, lakes, rivers, railroads, roads are from: <https://www.diva-gis.org/gdata>.
- Raster DHS covariates are from a different Measure DHS dataset. DHS program collects GPS data using Neighborhood calculations for raster data or Distance calculations using vector data. It uses centroids of 10 km for rural areas and 2 km for urban areas in calculations. See reports in: <https://dhsprogram.com/What-We-Do/GIS.cfm>.
- Raster DHS covariates include: populations indicators, nightlights, travel times, human footprint, proximity to borders and water, aridity, drought episodes, enhanced vegetation, growing season length, rainfall, temperature, etc.
- Other IPUMS-DHS contextual variables: battles, riots, cropland, pastureland, production, NVDI, geo GDP-economic activity, etc.

3.3. Indicators

The SEI website contains information for 14 education indicators on access and completion. Indicators' definitions are based on the **WIDE database** (<https://www.education-inequalities.org/>) which accounts for consistency while boosting N size of each education indicator. All education indicators are means at the community level (Table 3 and Figure 2 for details).

Table 3: Indicators - acronyms

| Type | Name | Acronym |
|---------------|------------------------------------|-------------------------|
| Access | Never been to school | c_edu0_prim |
| Completion | Prim completion rate | c_comp_prim |
| | Lowsec completion rate | c_comp_lowsec |
| | Upsec completion rate | c_comp_upsec |
| | Tert completion rate | c_comp_higher_2529_2yrs |
| Overage | Overage prim school attendance | c_ownership2plus |
| Out of school | Out of school rate at prim level | c_edu_out_pry |
| | Out of school rate at lowsec level | c_edu_out_lowsec |
| | Out of school rate at upsec level | c_edu_out_upsec |
| Transition | Transition from prim to lowsec | c_trans_prim |
| | Transition from lowsec to upsec | c_trans_lowsec |
| | Transition from upsec to HE | c_trans_upsec |
| Attainment | Mean years of education | c_eduyears_20 |
| Literacy | Youth literacy rate | c_literacy_1524 |

| category | name | description |
|---------------|---------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| Access | Never been to school | % of children aged 3-6 years above primary school entrance age who have never been to school. |
| Completion | Primary completion rate | % of children and young people aged 3-5 years above primary school graduation age who have completed primary school. |
| | Lower secondary completion rate | % of young people aged 3-5 years above lower secondary school graduation age who have completed lower secondary school. |
| | Upper secondary completion rate | % of young people aged 3-5 years above upper secondary school graduation age who have completed upper secondary school. |
| | Tertiary completion rate | % of people aged 25–29, who have completed at least two years of higher education. |
| Overage | Overage primary school attendance | % of children in primary school who are two years or more older than the official age for grade. |
| Out of school | Out of school rate at primary level | % of children of primary school age who are not in school. |
| | Out of school rate at lower secondary level | % of adolescents of lower secondary school age who are not in school. |
| | Out of school rate at upper secondary level | % of youth of upper secondary school age who are not in school. |
| Transition | Transition from primary to lower secondary | # of young people attending the first grade of lower secondary school as a percentage of those attending the final grade of primary school. |
| | Transition from lower to upper secondary | # of young people attending the first grade of upper secondary school as a percentage of those attending the final grade of lower secondary school. |
| | Transition from upper to higher education | # of young people attending the first grade of higher education as a percentage of those attending the final grade of upper secondary school. |
| Attainment | Mean years of education | Average number of years of schooling attained for the age group 20–24 years. |
| Literacy | Youth literacy rate | % of young women aged 15-24 who can read a simple sentence. |

Figure 2: Indicators descriptions

Table 2: Summary statistics of community educational outcomes (access, completion and overage) per country

| Country | Never been to school | | Completion primary | | Completion lower secondary | | Completion upper secondary | | Completion tertiary | | Overage | |
|---------------|----------------------|--------|--------------------|--------|----------------------------|--------|----------------------------|--------|---------------------|--------|---------|--------|
| | | | | | | | | | | | | |
| | mean | mean N | mean | mean N | mean | mean N | mean | mean N | mean | mean N | mean | mean N |
| Angola | 0.25 | 17.3 | 0.47 | 11.8 | 0.25 | 9.7 | 0.12 | 8.8 | 0.04 | 7.4 | | |
| Benin | 0.19 | 19.7 | 0.55 | 11.2 | 0.24 | 7.9 | 0.09 | 7.7 | 0.05 | 7.4 | | |
| Burkina Faso | 0.37 | 21.5 | 0.30 | 12.4 | 0.09 | 10.4 | 0.02 | 9.7 | 0.01 | 9.2 | | |
| Burundi | 0.10 | 21.6 | 0.53 | 13.8 | 0.25 | 11.3 | 0.10 | 10.9 | 0.02 | 9.6 | | |
| Cameroon | 0.09 | 15.6 | 0.70 | 12.0 | 0.31 | 10.2 | 0.10 | 9.5 | 0.07 | 8.6 | | |
| Chad | 0.48 | 27.9 | 0.25 | 16.7 | 0.12 | 11.6 | 0.06 | 9.4 | 0.02 | 9.4 | | |
| Côte d'Ivoire | 0.27 | 18.7 | 0.40 | 12.5 | 0.22 | 12.6 | 0.11 | 12.2 | 0.05 | 10.9 | | |
| D. R. Congo | 0.16 | 26.1 | 0.54 | 16.5 | 0.45 | 15.4 | 0.17 | 12.7 | 0.05 | 12.0 | | |
| Ethiopia | 0.19 | 17.3 | 0.53 | 10.6 | 0.34 | 8.3 | 0.32 | 8.9 | 0.17 | 8.9 | | |

Table 4: Moran I correlation (using power function, decay distance parameter = 1, and threshold distance = +inf)

| | Never been to school | Completion primary | Completion lower secondary | Completion upper secondary | Completion tertiary | Overage | Out of school primary | Out of school lower secondary |
|---------------|----------------------|--------------------|----------------------------|----------------------------|---------------------|---------|-----------------------|-------------------------------|
| Angola | 0.158 | 0.244 | 0.261 | 0.259 | 0.097 | 0.134 | 0.149 | 0.092 |
| Benin | 0.214 | 0.189 | 0.185 | 0.275 | 0.250 | 0.154 | 0.225 | 0.148 |
| Burkina Faso | 0.160 | 0.155 | 0.159 | 0.104 | 0.093 | 0.086 | 0.187 | 0.137 |
| Burundi | 0.125 | 0.144 | 0.116 | 0.160 | 0.108 | 0.100 | 0.155 | 0.134 |
| Cameroon | 0.258 | 0.356 | 0.252 | 0.234 | 0.164 | 0.366 | 0.310 | 0.188 |
| Chad | 0.287 | 0.334 | 0.320 | 0.336 | 0.340 | 0.042 | 0.296 | 0.299 |
| Côte d'Ivoire | 0.097 | 0.114 | 0.107 | 0.124 | 0.139 | 0.106 | 0.088 | 0.042 |
| D. R. Congo | 0.102 | 0.184 | 0.221 | 0.367 | 0.254 | 0.145 | 0.069 | 0.071 |
| Ethiopia | 0.117 | 0.131 | 0.189 | 0.213 | 0.224 | 0.092 | 0.120 | 0.075 |

Summary statistics EI

Table 5: Moran I correlation (using power function, decay distance parameter = 1, and threshold distance = average distance community / 2)

| | Never been to school | Completion primary | Completion lower secondary | Completion upper secondary | Completion tertiary | Overage | Out of school primary | Out of school lower secondary |
|---------------|----------------------|--------------------|----------------------------|----------------------------|---------------------|---------|-----------------------|-------------------------------|
| Angola | 0.247 | 0.324 | 0.333 | 0.320 | 0.125 | 0.169 | 0.221 | 0.144 |
| Benin | 0.401 | 0.299 | 0.247 | 0.315 | 0.276 | 0.229 | 0.411 | 0.280 |
| Burkina Faso | | | | | | | | |
| Burundi | 0.289 | 0.306 | 0.208 | 0.249 | 0.162 | 0.192 | 0.339 | 0.306 |
| Cameroon | 0.462 | 0.573 | 0.354 | 0.291 | 0.204 | 0.562 | 0.545 | 0.346 |
| Chad | 0.464 | 0.465 | 0.388 | 0.380 | 0.378 | 0.046 | 0.470 | 0.493 |
| Côte d'Ivoire | 0.195 | 0.165 | 0.157 | 0.178 | 0.193 | 0.165 | 0.142 | 0.093 |
| D. R. Congo | | | | | | | | |
| Ethiopia | 0.235 | 0.186 | 0.246 | 0.270 | 0.270 | 0.153 | 0.228 | 0.137 |

Moran I - version 1

Table 7: Empirical Bayes Kriging (EBK) mean prediction statistics

| | Never been to school | Completion primary | Completion lower secondary | Completion upper secondary | Completion tertiary | Overage | Out of school primary | Out of school lower secondary |
|--------------|----------------------|--------------------|----------------------------|----------------------------|---------------------|---------|-----------------------|-------------------------------|
| Angola | | | | | | | | |
| RMSE | 0.1909 | 0.2345 | 0.2007 | 0.1479 | 0.1106 | 0.1709 | 0.1923 | 0.2064 |
| RMSE SD | 0.9830 | 0.9848 | 1.0208 | 1.1985 | 1.8837 | 1.0356 | 0.1923 | 0.9988 |
| ASE | 0.1883 | 0.2418 | 0.2072 | 0.1355 | 0.0748 | 0.1624 | 0.1936 | 0.1972 |
| Benin | | | | | | | | |
| RMSE | 0.1587 | 0.2209 | 0.2007 | 0.1210 | 0.0873 | 0.1333 | 0.1380 | 0.1981 |
| RMSE SD | 0.9839 | 0.9944 | 1.0000 | 1.4000 | 2.1837 | 0.9963 | 1.0181 | 0.9945 |
| ASE | 0.1565 | 0.2286 | 0.2077 | 0.1079 | 0.0704 | 0.1324 | 0.1564 | 0.1970 |
| Burkina Faso | | | | | | | | |
| RMSE | 0.2037 | 0.2059 | 0.1230 | 0.0643 | 0.0460 | 0.1720 | 0.1938 | 0.2089 |
| RMSE SD | 1.0139 | 0.9979 | 1.0213 | 1.0485 | 2.1530 | 1.0234 | 0.9913 | 0.9877 |
| ASE | 0.2022 | 0.2221 | 0.1184 | 0.0487 | 0.0335 | 0.1645 | 0.1975 | 0.2153 |
| Burundi | | | | | | | | |
| RMSE | 0.0873 | 0.1757 | 0.1726 | 0.1119 | 0.0650 | 0.1288 | 0.1079 | 0.1593 |
| RMSE SD | 0.9982 | 0.9870 | 0.9887 | 1.0780 | 1.7904 | 0.9849 | 0.9751 | 1.0087 |
| ASE | 0.0864 | 0.1786 | 0.1745 | 0.1066 | 0.0456 | 0.1309 | 0.1104 | 0.1615 |

Moran I - version 2

Interpolation statistics

Figure 3: Example. Tables - SEI website

3.4. Tables

Under the resources tab, the SEI website contains several tables, which can be downloaded as *.csvs for further use (see: Figure 3). Specifically, tables included are:

- List of countries and N size.
- Summary statistics for each country EI with mean value and mean community N used for calculations.
- Spatial autocorrelation (for error term: *Moran I*) obtained under 3 alternatives.
 1. Using power function: decay distance parameter $\delta = 1$ and threshold distance $d = \infty$.
 2. Using power function: $\delta = 1$ and $d = \text{mean distance communities} / 2$.
 3. Using exponential function: $\delta = 1$ and threshold distance $d = \infty$.
- Also, *Moran I* tables includes mean, minimum and maximum distance across communities.
- Empirical Bayes Kriging (EBK) mean prediction statistics for assessing the validity of interpolation maps. Statistics included are: Root Mean Squared Error (RMSE), Root Mean Squared Error standardised (RMSE SD), and Average Squared Error (ASE).

3.5. Maps

The SEI website contains three types of maps:

1. Distribution and local correlation → **maps type 1**.
2. Mean predicted – Interpolation → **maps type 2**.

3. Mean predicted – Interpolation. Inequality ratios → **maps type 3**.

Within each map type, there are different versions. Taking into account the current number countries and education indicators (EI) included so far in the SEI website, the total number of maps is shown in Table 4.

Table 4: Number of maps

| Type | # of countries | # of EI | # of maps | Total # of maps |
|----------|----------------|---------|----------------|-----------------|
| 1 | 29 | 14 | 4 | 1,624 |
| 2 | 29 | 14 | 6 | 2,436 |
| 3 | 29 | 14 | 5 (× 4 ratios) | 8,120 |

3.5.1. Distribution and local correlation

The SEI website begins by showing the distribution of education indicators (EI) for each country using quartiles and then plotting *Moran I* local correlations across each country map. Specifically, there are four maps versions for each EI:

1. Quartiles (natural breaks, Jenks).
2. Quartiles with water areas and railway lines.
3. Hot and cold spots (local correlation) with water areas and railway lines.
4. Hot and cold spots (local correlation) with water areas and railway lines. Each hot-cold spot has a buffer of 5 km.

Figure 4 contains an example for one country/indicator.

3.5.2. Mean predicted – Interpolation

In this section, the aim of the maps is to provide an smoothing surface of the prediction for each EI. Hence, the chosen layers used, to pin down and identify pockets of education disadvantages, are administrative borders. Two borders, at administrative levels 1 and 2, are employed as a way to allow stakeholders to identify particular ad hoc areas for educational targeting. Because communities' sampling points are smoothed and obtained through Empirical Bayesian Kriging (see section 2.4), statistics related to the overall prediction are also provided (see: Figure 3) but, importantly each map prediction can be contrasted with its standard deviation (SD), showing the accuracy of the prediction also included in maps.

In particular, mapping for this stage contains six maps:

1. Empirical Bayesian Kriging (EBK) education indicator (EI) predicted mean.
2. EBK EI predicted mean standard deviation.
3. EBK EI predicted mean with administrative boundary 1.
4. EBK EI predicted mean standard deviation with administrative boundary 1.
5. EBK EI predicted mean with administrative boundary 2.
6. EBK EI predicted mean standard deviation with administrative boundary 2.

Figure 5, as an example, uses Nigeria and overage at primary level as an indicative education indicator. The six maps, one prediction and one for the precision of the prediction are shown, under different administrative boundaries.

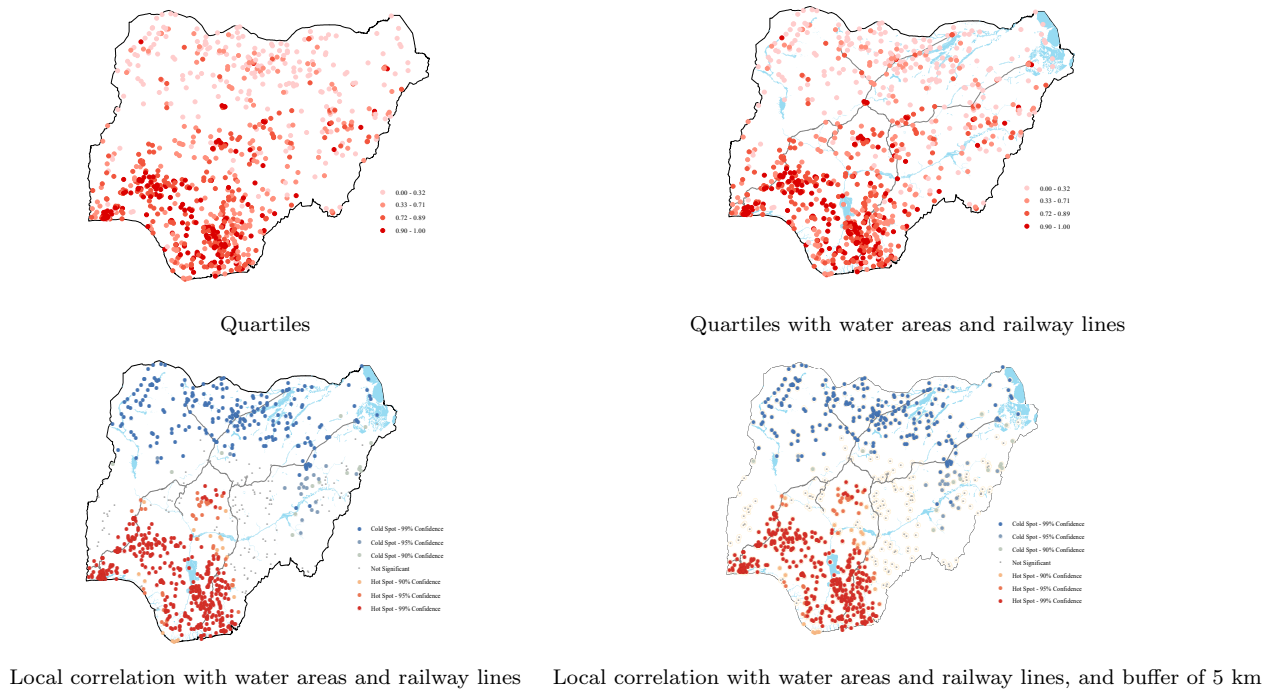


Figure 4: Maps type 1 - lower secondary completion, Nigeria

3.5.3. Mean predicted – Interpolation. Inequality ratios

The last set of maps included are on inequality ratios for EI. *Thus far*, the SEI website focus is on four ratios: gender, wealth, early marriage and empowerment, with ratios for education disadvantage calculated as:

1. Gender (**g**). Ratio = (female / male).
2. Wealth (**w**). Ratio = (poor household / non-poor household), where poor household is defined at those families whose wealth is on the bottom three quartiles and non-poor households in the top two quartiles.
3. Early marriage (**e2**). Ratio = (women marriage between age 10-17 / women married age 18 or above).
4. Empowerment (**vf**). Ratio = (women empowered on family visits decision / women non-empowered on family visits decision), where empowered is defined if decision is taken by either women or jointly women and husband/partner, and non-empowered if decision is taken by husband/partner alone.

Here, note that instead of using a continuous version of the interpolation, a quintile division of the predicted mean of the EI is used. The reason is that ratios are more easily identified within fixed intervals and related targeting on given sources of inequalities is, hence, more efficient. There are five maps for each ratio.

1. EBK predicted mean in quintiles.
2. EBK predicted mean in quintiles with community location observations.

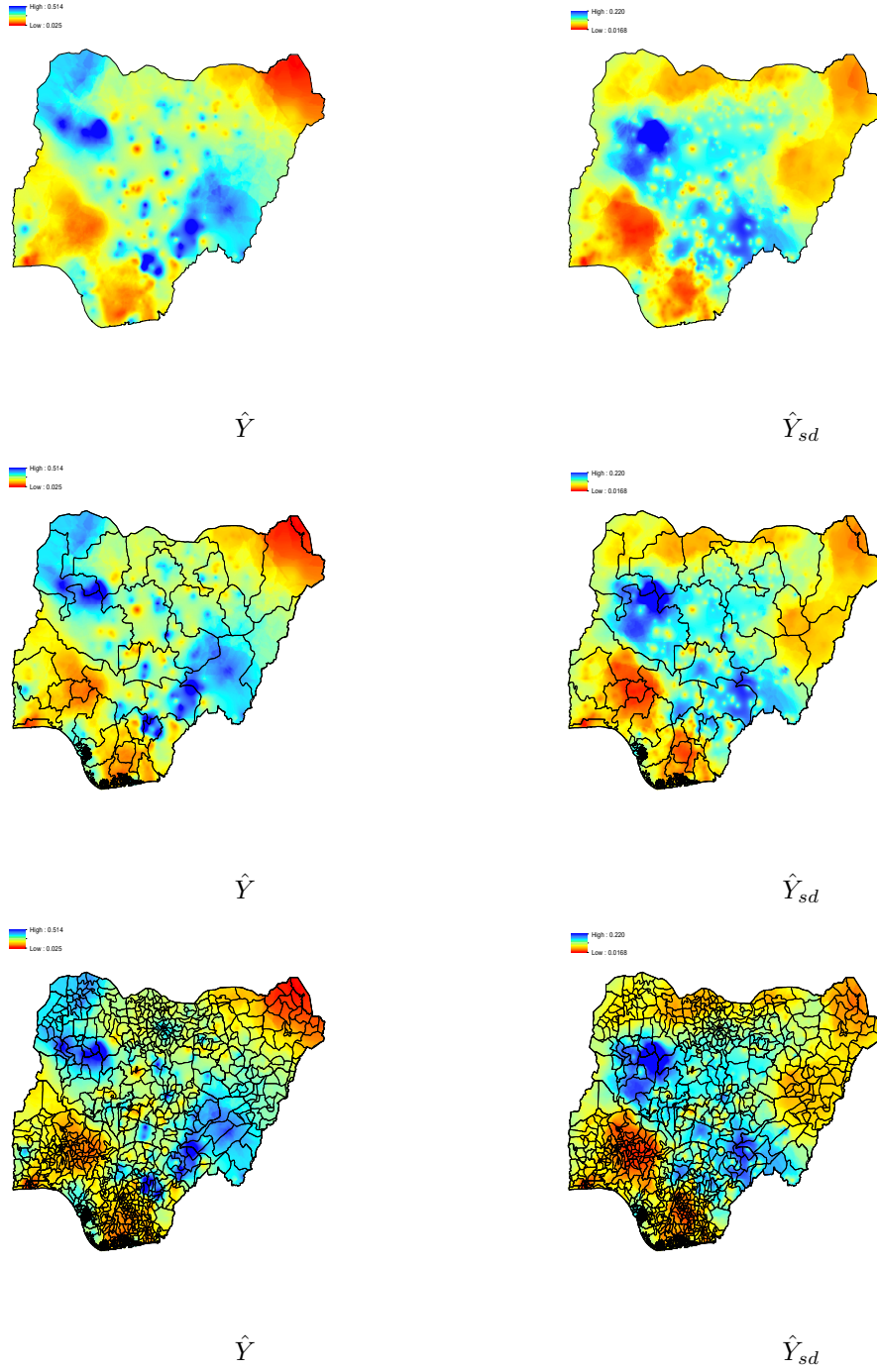
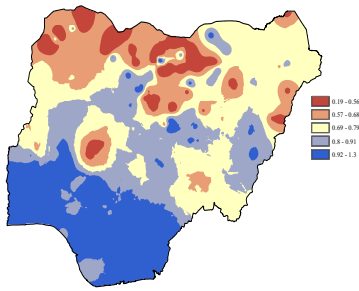


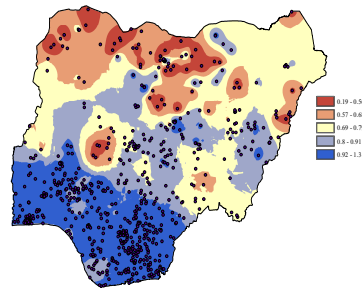
Figure 5: Maps type 2 - primary overage, Nigeria

3. EBK predicted mean in quintiles with community location observations with roads.
4. EBK predicted mean in quintiles with administrative boundary 1.
5. EBK predicted mean in quintiles with administrative boundary 2.

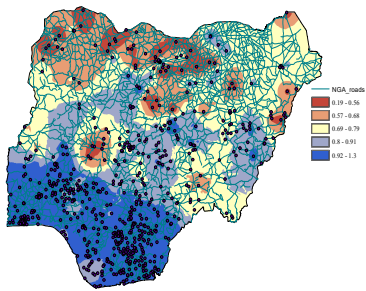
Giving the number of ratios and EI, maps are numerous (see Table 4). Figures 6 to 9 display the ratio for the indicator lower secondary completion (Nigeria) as an example.



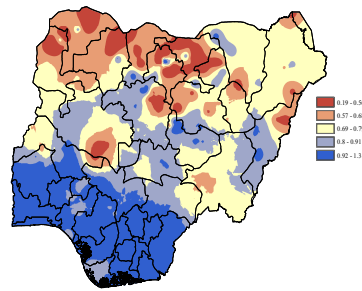
quintiles



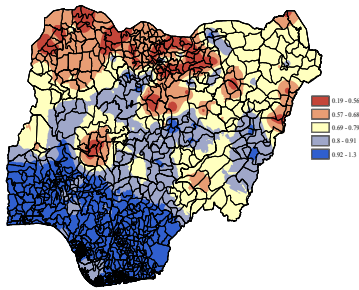
quintiles with community locations



quintiles with community locations

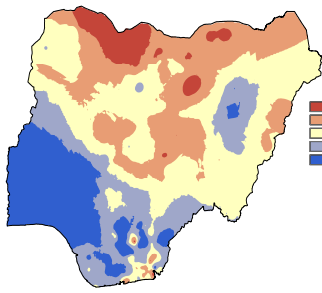


quintiles with admin boundaries 1

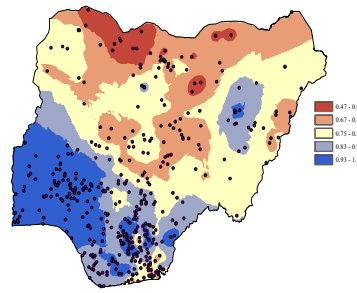


quintiles with admin boundaries 2

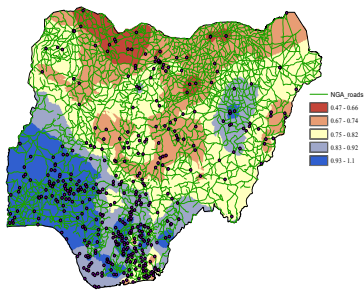
Figure 6: Maps type 3 (**gender**) - lower secondary completion, Nigeria



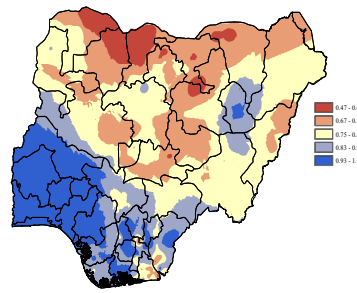
quintiles



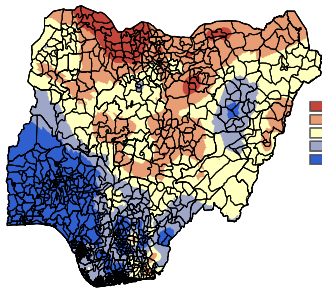
quintiles with community locations



quintiles with community locations

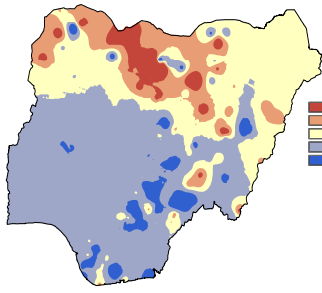


quintiles with admin boundaries 1

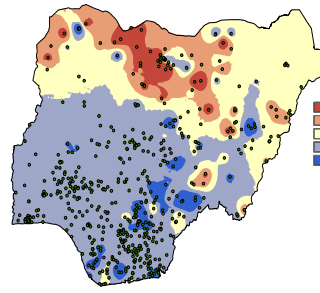


quintiles with admin boundaries 2

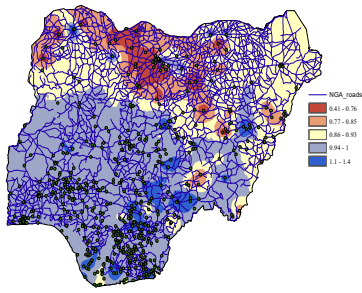
Figure 7: Maps type 3 (**wealth**) - lower secondary completion, Nigeria



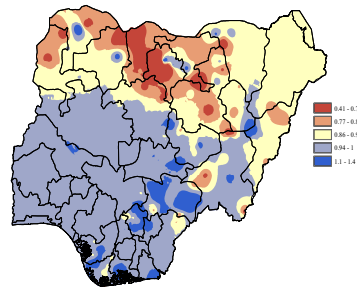
quintiles



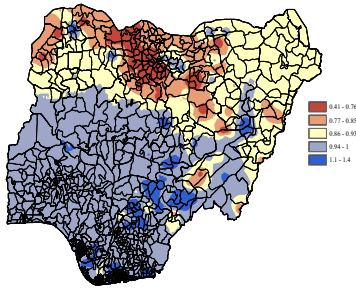
quintiles with community locations



quintiles with community locations

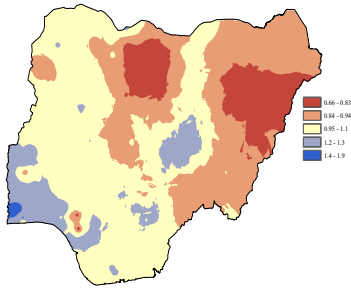


quintiles with admin boundaries 1

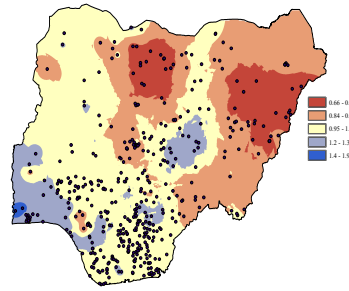


quintiles with admin boundaries 2

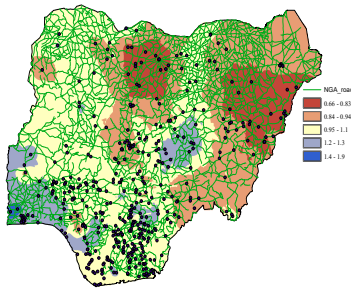
Figure 8: Maps type 3 (**early marriage**) - lower secondary completion, Nigeria



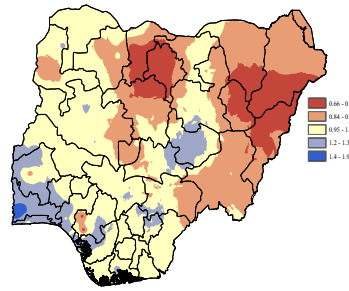
quintiles



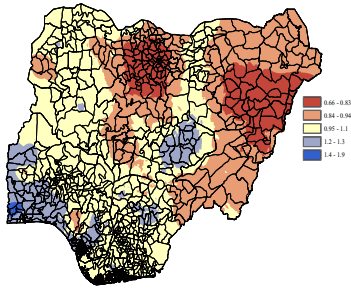
quintiles with community locations



quintiles with community locations



quintiles with admin boundaries 1



quintiles with admin boundaries 2

Figure 9: Maps type 3 (**empowerment**) - lower secondary completion, Nigeria

3.6. Files

Stakeholders' involvement from a spatial perspective, be it either at a country level or international organisations monitoring SDG4, is a key objective of the SEI website. In consequence, this involvement is facilitated by providing files used for maps' construction so users can rely on them for further analysis. For each coun-

try/indicator, two types of files are provided:

1. Excel files with coordinates/XY location for each community and the mean value of the EI.
2. Geo located files with estimations and layers in the form of ArcGIS files (*.mxd) which users can further manipulate and tailor to their interests using the ArcGIS Desktop software.⁶

The access to these files permits users to engage in showcased analysis beyond what is offered by the maps in the SEI website and, importantly, Excel files can be used relying on other free software such as R software.⁷

The number of files for additional use are linked to the maps provided (Table 5). In the website, in the process of downloading maps users can get these files.

Table 5: Number of Excel and ArcGIS files

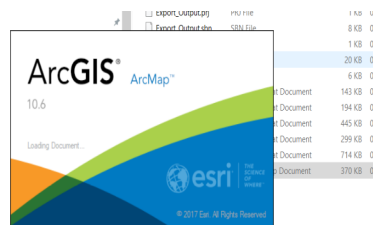
| Map type | # of countries | # of EI | # of Excel files | # of ArgGIS files |
|----------------|----------------|---------|------------------|-------------------|
| 1 and 2 | 29 | 14 | 406 | 406 |
| 3 | 29 | 14 | 406 x 4 | 406 x 4 |

| | A | B | C | D | E | F |
|----|---------|--------|------------|-----|-------|-------|
| 1 | country | crural | c_rsdg_pgm | _ID | _CX | _CY |
| 2 | Nigeria | 1 | 0.1429 | 1 | 8.087 | 6.902 |
| 3 | Nigeria | 1 | 0.0000 | 2 | 7.877 | 7.854 |
| 4 | Nigeria | 1 | 0.0208 | 3 | 7.940 | 7.781 |
| 5 | Nigeria | 1 | 0.1111 | 4 | 8.197 | 7.177 |
| 6 | Nigeria | 1 | 0.0355 | 5 | 8.164 | 7.149 |
| 7 | Nigeria | 1 | 0.2571 | 6 | 8.921 | 7.457 |
| 8 | Nigeria | 1 | 0.1500 | 7 | 9.005 | 7.445 |
| 9 | Nigeria | 1 | 0.0952 | 8 | 8.673 | 7.246 |
| 10 | Nigeria | 0 | 0.0000 | 9 | 8.977 | 7.334 |
| 11 | Nigeria | 1 | 0.0952 | 10 | 8.209 | 7.297 |
| 12 | Nigeria | 1 | 0.1379 | 11 | 9.678 | 7.188 |
| 13 | Nigeria | 1 | 0.1667 | 12 | 9.682 | 7.293 |
| 14 | Nigeria | 0 | 0.0000 | 13 | 8.553 | 7.727 |
| 15 | Nigeria | 1 | 0.1018 | 14 | 8.307 | 6.956 |
| 16 | Nigeria | 1 | 0.0000 | 15 | 7.685 | 7.119 |
| 17 | Nigeria | 0 | 0.1000 | 16 | 8.434 | 6.848 |
| 18 | Nigeria | 1 | 0.0000 | 17 | 7.813 | 7.178 |
| 19 | Nigeria | 1 | 0.0000 | 18 | 7.796 | 7.122 |
| 20 | Nigeria | 1 | 0.1111 | 19 | 8.089 | 7.103 |
| 21 | Nigeria | 1 | 0.1176 | 20 | 7.967 | 7.481 |
| 22 | Nigeria | 0 | 0.0667 | 21 | 8.146 | 7.191 |
| 23 | Nigeria | 1 | 0.4000 | 22 | 9.538 | 7.782 |
| 24 | Nigeria | 1 | 0.2273 | 23 | 9.621 | 7.731 |
| 25 | Nigeria | 1 | 0.2368 | 24 | 9.075 | 6.714 |
| 26 | Nigeria | 1 | 0.1111 | 25 | 6.802 | 8.633 |
| 27 | Nigeria | 0 | 0.0556 | 26 | 6.947 | 8.483 |
| 28 | Nigeria | 1 | 0.0000 | 27 | 7.379 | 9.056 |
| 29 | Nigeria | 1 | 0.0000 | 28 | 7.422 | 9.025 |

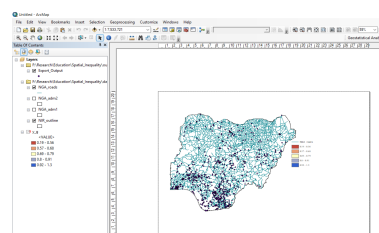
Excel file

| Name | Type | Size |
|-------------------|------------------------|--------|
| Export_Output.cpg | CPS File | 1 KB |
| Export_Output.dbf | DBF File | 253 KB |
| Export_Output.prj | PRJ File | 1 KB |
| Export_Output.sbn | SBN File | 8 KB |
| Export_Output.sbx | SBX File | 1 KB |
| Export_Output.shp | SHP File | 20 KB |
| Export_Output.shx | SHX File | 6 KB |
| map1 | Adobe Acrobat Document | 143 KB |
| map1_v2 | Adobe Acrobat Document | 194 KB |
| map1_v3 | Adobe Acrobat Document | 445 KB |
| map2 | Adobe Acrobat Document | 299 KB |
| map3 | Adobe Acrobat Document | 714 KB |
| Untitled | ArcGIS ArcMap Document | 370 KB |

ArcGIS



ArcGIS



ArcGIS

Figure 10: Files

⁶<http://desktop.arcgis.com/en/arcmap/>.

⁷See, for instance: http://www.econ.uiuc.edu/~lab/workshop/Spatial_in_R.html and free R packages: `spdep` and `maptools`.

4. SEI website - what is online

4.1. Overview - menu and resources

The front page of the SEI website www.spatial-education-inequalities.com contains the main menu, i.e., a **Country** and **Indicator** search feature (Figure 11).

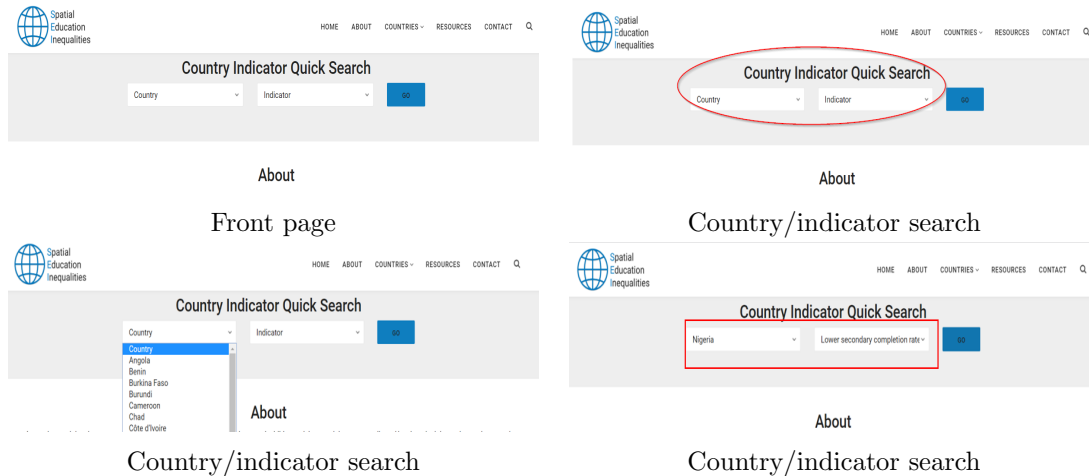


Figure 11: SEI website, front page

In the *resources tab*, in addition to methodological notes, tables with relevant statistics are included (Figure 12).

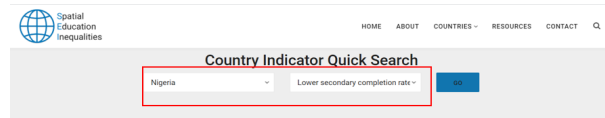


Figure 12: SEI website, resources

4.2. Menu - Maps

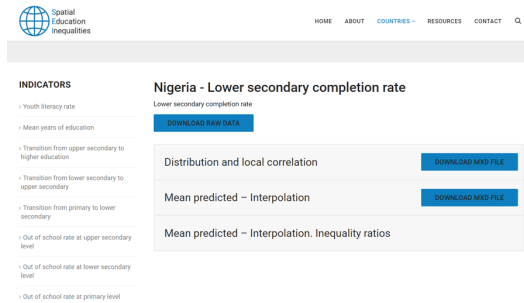
Once users have chosen the country/indicator combination, the three types of maps menu appears -i.e., distribution, interpolation and interpolation for ratios. On

top, there is the option to download the Excel raw data with XY coordinates and community mean value of the given EI.

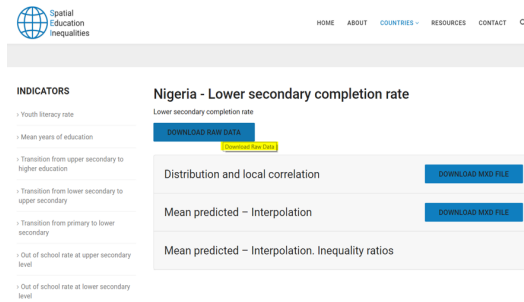


About

Country/indicator



Maps menu



Maps menu, Excel downloadable file

Figure 13: SEI website, maps menu

After users click on one of the three tabs of the maps' menu, a description of what it is included appears with the array of maps underneath. There are two options below each displayed maps, either to download it or to expand it onto the browser. The sub-sections below display how the array of maps are presented in the website.

4.2.1. Maps - type 1, distribution

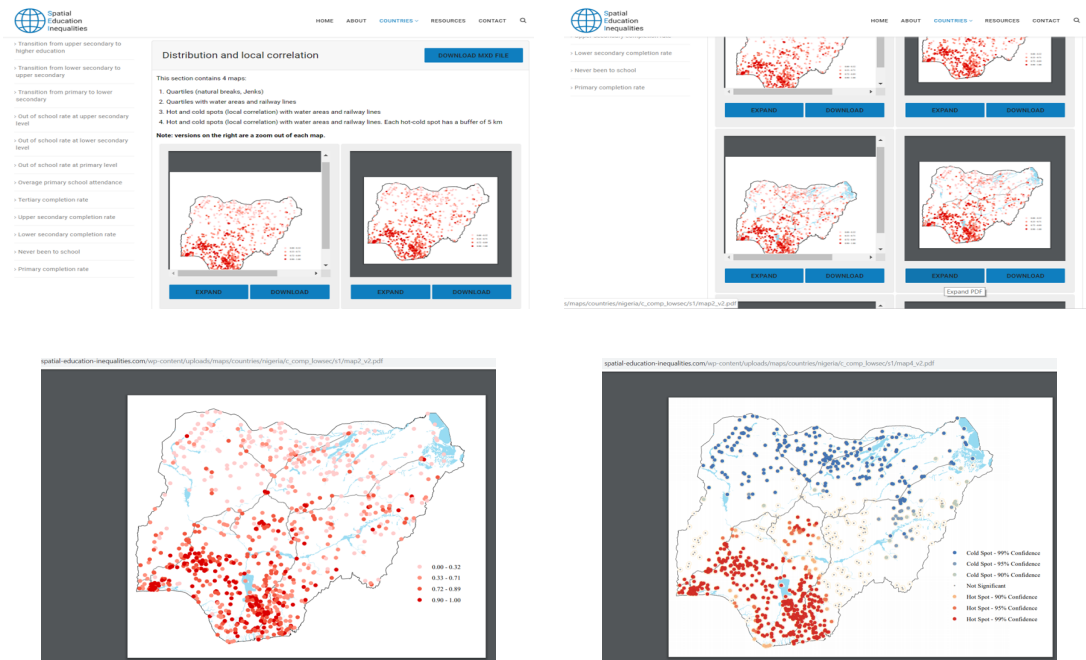


Figure 14: Maps type 1

4.2.2. Maps - type 2, interpolation

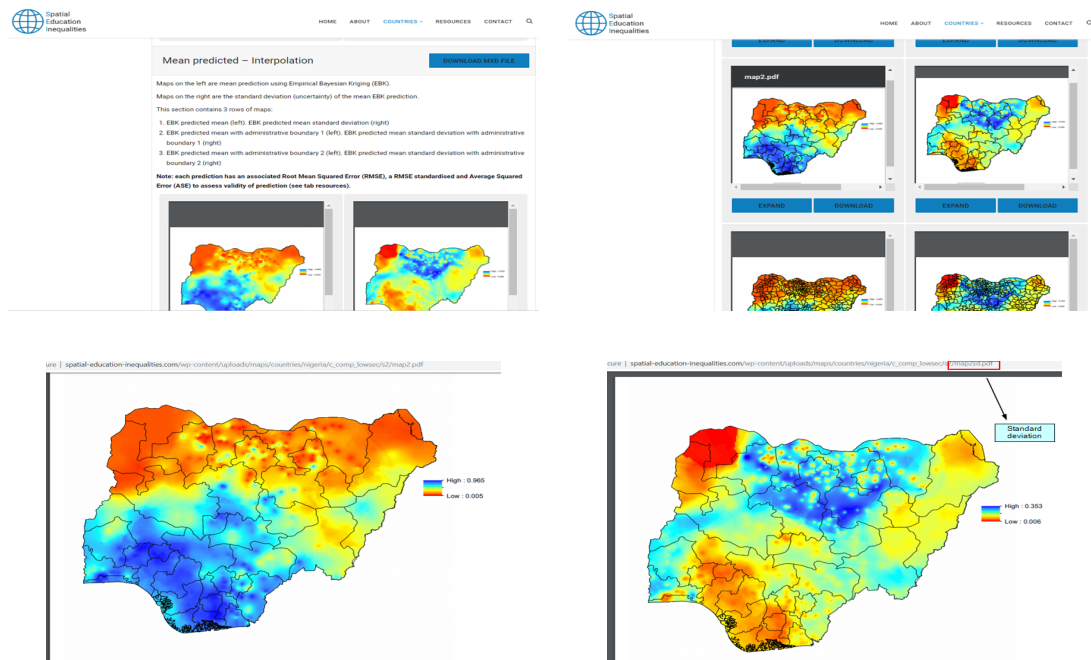


Figure 15: Maps type 2

4.2.3. Maps - type 3, inequality ratios

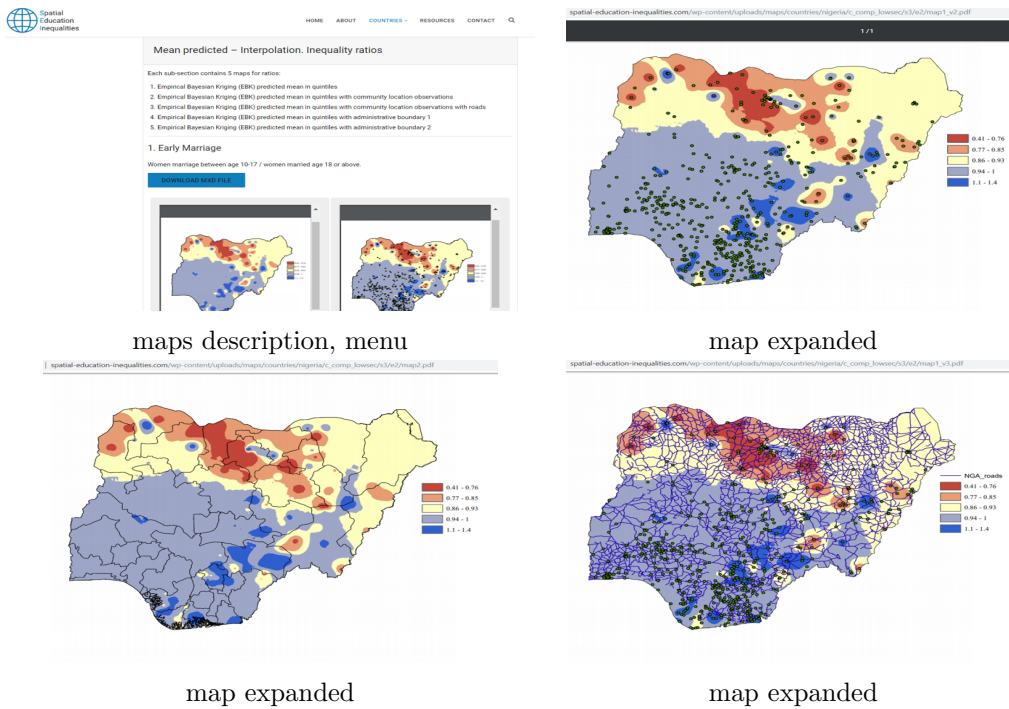


Figure 16: Maps type 3, early marriage

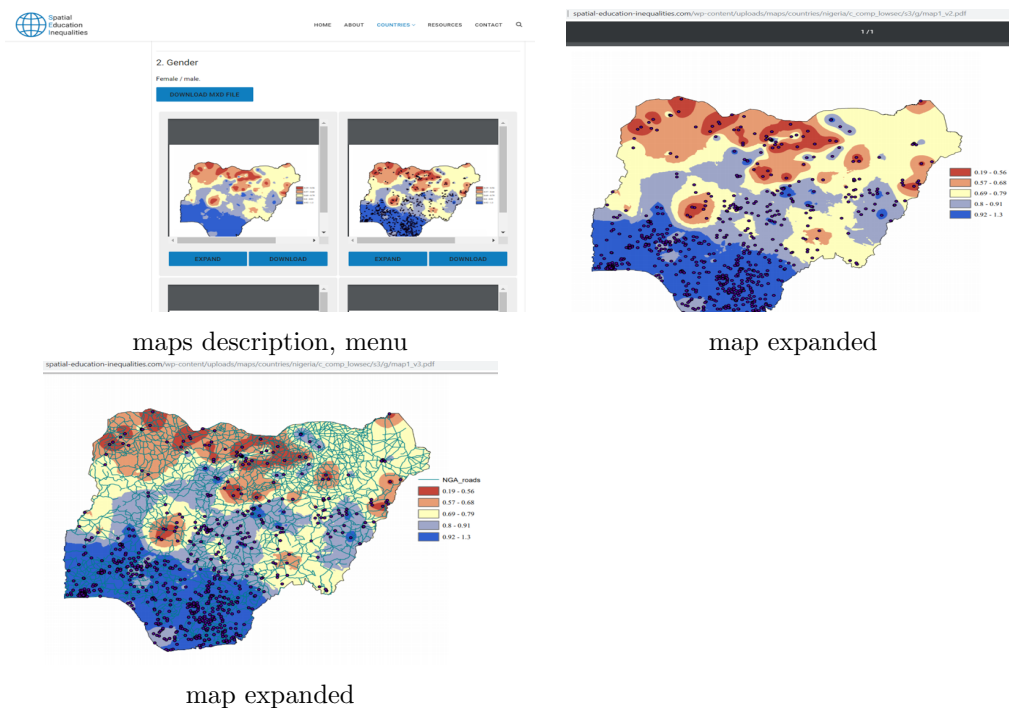
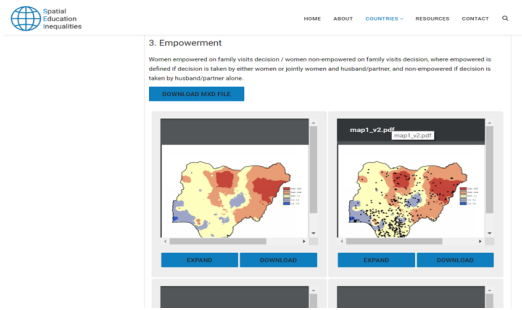
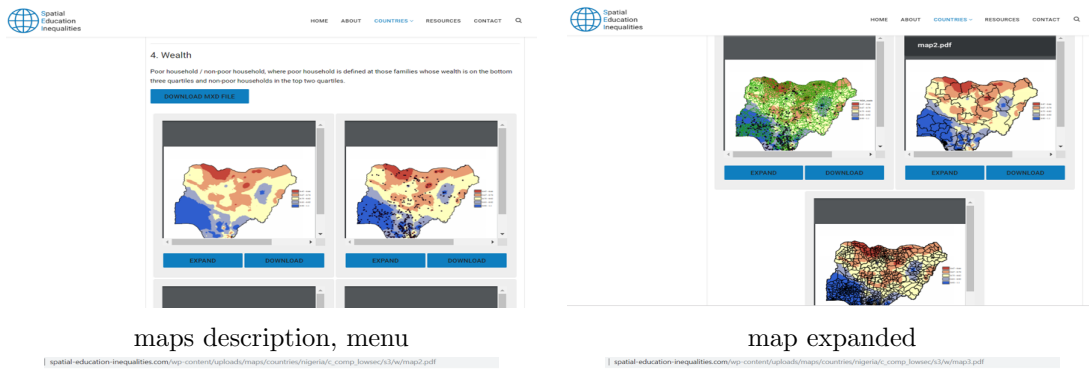


Figure 17: Maps type 3, gender



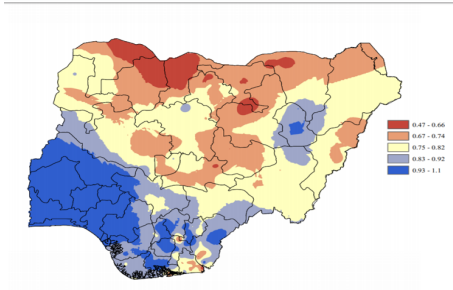
maps description, menu

Figure 18: Maps type 3, **empowerment**

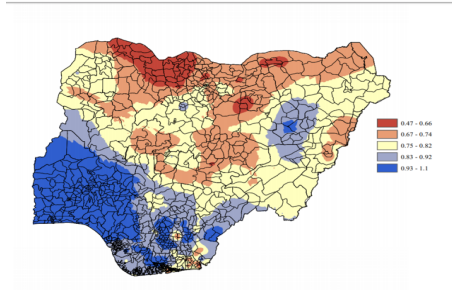


maps description, menu

map expanded



map expanded



map expanded

Figure 19: Maps type 3, **wealth**

5. Conclusions

This background note highlighted the motivation behind the construction of the SEI website, its technical details and contents of the site, as well as what is currently online guiding the user across the site different components.

The SEI website was designed to fill the gap underpinning the push for more data-driven analysis in the monitoring the SDG4 agenda using current available DHS surveys from a spatial perspective. In doing so, this new (and work in progress) monitoring tool makes possible for stakeholders the exploration a new dimension of education inequality, that is, spatial inequality within a country, but at a more refined and granular level. The long term objective of the SEI website is to initiate

and generate a wider discussion on the role of “space and education inequality” within SDG4 and its cross synergies and implications for other SDGs.